Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 21 Issue 16, December 2019

Real Time Sign Detection And Motion Prediction System for ISL Using Deep Learning.

Shraddha Bhaware[#1],Prof. M. S. Wakode[*2],

Department of Computer Engineering, PICT, Pune, India
[1]*shraddha.bhaware@gmail.com*
[2]*mswakode@pict.edu*

***Abstract*—**Hand gestures are the nonverbal method of communication used along with verbal statement for transferring the information. Ideally, a mechanism in which the hand gesture plays a vital role for information interchange is called as "Sign Language". In this sign language individually alphabet of the English vocabulary is assigned with a unique sign. The main objective of this project is to develop a system to support verbally challenged people. In the proposed system, the novel mechanism using neural networks has been proposed for real time dynamic gesture recognition for Indian English sign language dataset which accepts live video stream as input and displays the predicted text output for the trained detected sign. The live stream is divided into many fragments for feature extraction and the person performing the gesture is identified as a subject. After this, the background is disregarded and detection of the hand is performed in every frame of the given input live video. As soon as the hand sign is detected, the movement of hand is fetched and corresponding text is displayed. This motion data is then accumulated from a live video stream using open-cv and Machine Learning model with CNN and represented in the form of interval-valued data synthesis is performed. A suitable classifier is generated based on these statistics. Testing can be completed to obtain the efficiency of the given system. The given testing input is checked to be within the range of given intermission values and then confirmed to be a certain gesture.

Keywords*-Convolutional Neural Network(CNN), Sign Detection, Machine learning,Gesture Recognition, Image Processing.*

I. INTRODUCTION

The sign language (known as the signed language) is a language that practice the visual-guide modality to tell the meaning of spoken statement by any individual person. Sign languages are full established natural languages with their own grammar and formation. This means that sign languages are not universal and they are not commonly understandable, although there are also striking resemblances among sign languages.

On the other hand, from the human perception the most basic and natural way to co-operate with the computer is through talking and hand gesture interface. Thus, the research of sign language and gesture acknowledgment is possible to provide a shift paradigm from out-of-date point of click, user interface to a natural language dialogue and vocalized command-based interface. So, in Human-Computer

ISSN: 0374-8588
Volume 21 Issue 16, December 2019

Interaction(HCI), sign language recognition(SLR) being considered as an essential work area, has concerned more and additional awareness to researchers to show interest in HCI society.

The whole idea behind this dissertation is to detect and predict different signs from a set of predefined signs of Indian sign language. The comparative study is carried out on the basis of differently generated data is passed to the same machine learning algorithm. The basic idea is to build a real time system to detect human understandable sign gestures using a camera based live video input. The model will be trained over a dataset which is created using signers and five to eight different classes. The CNN model were trained and tested for the sake of optimal result and simplicity in design and usability. This model is a CNN network i.e. model which is trained on pre-processed sign image dataset without any complex pre-processing wherein we can directly input live video stream with camera by a signer and fed it to the network.

M. R. Abid, et al. [9] presented a system which is based on smart home interactivity application. The system proposed is known as dynamic sign language recognition (DSLR). It states that their DSLR system is able to rule out ungrammatical sentence and showed a performance accuracy of nearly 98%. P. V. V Kishore, et al. [10] proposed to implement Indian sign language which converts sign into words or sentences.The system showed an accuracy of 91 %. V. D. Edke, et al. [11] presented a paper which audits video content investigation from the different circumstances into issue form. J. Farooq and M. B. Ali [12] argued that hand signal acknowledgment is a characteristic and instinctive way to associate with the PC, since collaborations with the PC can be expanded through multidimensional utilization of hand signals as contrast with other info strategies.The paper implements mainly three algorithms Convex –Hull, K-Curvature, Curvature of Perimeter. Guillaume Plouffe and Ana-Maria Cretu [13] introduced a paper which improves natural gesture user interface using data collected by a Kinect sensor. A tale technique is proposed to increase the time of filtering. J. Ravikiran, et al. [14], proposed a system which can recognize static hand gesture. The system concentrates on the number of fingers open and interpret it into an American Sign Language. Z. A. Ansari and G. Harit [15], state that using Microsoft's Kinect camera can give reading signs a better approach. The system has a prediction accuracy of above 90%. Hassan et al. [16] presented a solution for Hausa Sign Language. They have used Fourier descriptor implementing Particle Swarm Optimization to optimize the descriptor for extracting features from datasets. The prediction results accuracy was above 90%. A. Jarman, et al. [17] presented a novel algorithm to intercept Bengali sign language which can recognize almost 46 different hand gestures including vowels, consonant, numerical words. They have used multi layered feed-forward neural network with back-propagation. Around 2300 images were used and the accuracy reached was near 88%. S. Konwar, et al. [18] proposed a similar approach as discussed earlier. They have used HSV color model which can easily detect human skin. For feature extraction the authors proposes the use of edge detection algorithms and then applying a morphological operation to predict a given sign.

II. LITERATURE SURVEY

NO.	REFERENCE PAPER	DATASET AND TECHNIQUE	GAP
1	Indian sign language recognition using SVM [1]	Dataset:- Synthetic Dataset Technique:- Support vector classifier technique	The system worked on image capture mechanism which is not a feasible real time usable technique for the systems require live stream recognition
2	Indian sign language recognition system [2]	Dataset:-Publically available black background dataset Technique:- Support vector machine with Neural Network	The system worked on image capture mechanism which is not a feasible real time usable technique as such systems require live stream recognition.
3	Indian sign language recognition using neural networks and kNNclassifiers [3]	Dataset:- Custom synthetic dataset Technique:- KNN with Neural Networks	This system did not have the dynamic motion oriented recognition and did not have the real time live stream recognition. Only 10 numerical signs from 0-9 were recognised.
4	3D convolutional neural networks for dynamic sign language recognition [4]	Dataset:- ChaLearn dataset Technique:- 3D Convolutional Neural Networks	As this system worked on video input process, it did not have the real time live stream recognition. Also the dataset used is not an Indian dataset.
5	Dynamic hand gesture recognition using vision-based approach for human–computer interaction [5]	Dataset:- NITS hand gesture database IV Technique:- ANN, SVM, kNN	The proposed system could work only on the image dataset for numeric signs, alphabetic signs and arithmetic operators
6	Selfie video based continuous Indian sign language recognition system [6]	Dataset:- Synthetic smartphone captured dataset Technique:- Minimum distance classifier and Artificial Neural Network	The input video from the smartphone is checked for minimum distance, but there are many signs which pictorially depict same kind of gesture with minor pose change, such gestures won't be recognized accurately and the maximum accuracy of this system was 85% to 90%.
7	Real-time Indian sign language(ISL) recognition [7]	Dataset:- Full sleeve worn captured custom dataset Technique:-kNearest Neighbours algorithm and Hidden Markov Model	The proposed system works only on numbers and alphabets which is a very traditional dataset, so this system does not work to determine the emotion of words (actions).
8	Dynamic gesture recognition based on MEMP network [8]	Dataset:- LSA64, SKIG and Chalearn 2016 datasets Technique:- 3D CNN and ConvLSTM	The input is the gesture video which does notwork in real time scenario for live webcam gesture recognition.

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 16, December 2019

III. PROPOSED SYSTEM APPROACH

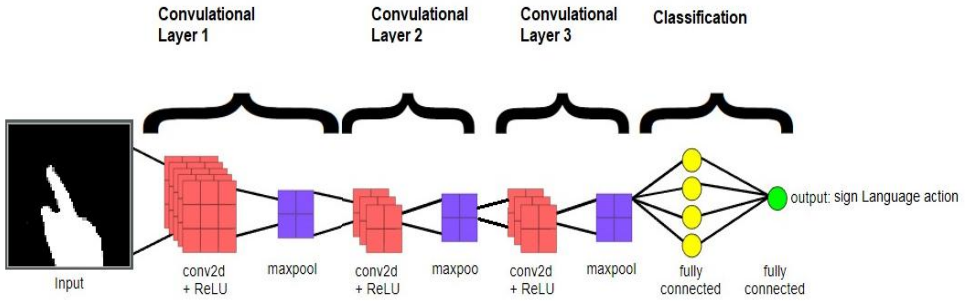


Fig.Proposed System

Indian sign language dataset is used to train the system,it consists of more than 9000 images of sign language actions, 1,100 images for each of the hand gestures. These were recorded from camera and then processed. The gestures include words like “Summer”, “Understand”, “Hello”, “Winter”, “Okay”, “Best”, “Like”, “Little”. The images are gray-scale with resolution of 64x64 pixel.

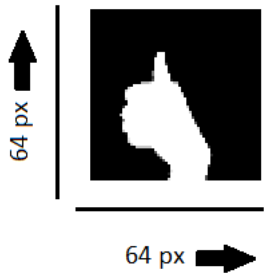


Fig.Sample Input Image

The dataset is divided into two sets as follows:

• Training dataset: This dataset contains words in which each word contains1000 images individually.

• Testing dataset: This is a subset of dataset used for training. This subset of dataset is used for validation testing.

In the proposed approach, the first stage and important stage is the object detection from extracted frames of video. The target of this stage is to detect hand substances in the digital images or videos. The most common problem is unstable brightness, noise, poor resolution and contrast. The better setting and camera devices can effectively advance these problems. The subsequent stage is object recognition basically hand sign recognition. The detected hand objects are recognized to identify the gesticulations. The image processing of frame consist of following steps:

A. *Segmentation:* The live stream is segmented into frames. The segmentation is the procedure of segregating a digital image into numerous segments or frames. The subdivision is to simplify or change the representation of an image into more communicative and easier to analyze the data.

B. *Binarization:* All the grey scale images are binarized with the assistance of algorithm. The algorithms should work well for images with intricate surrounding background. Here key frame extraction is performed on the basis of a pause to recognize the essential frames needed in order to categorize the given hand gesture stream.

C. *Thresholding:* Thresholding is to further-simplify realistic graphical data for analysis. First, you may convert to grey-scale, but then you have to consider that grayscale still has at lowest 255 values. In this proposed system with help of thresholding at the most basic level, is convert everything to white or black, based on a threshold value given. In this system the threshold to be 125 (out of 255), then everything that was 125 and under would be rehabilitated to 0, or black, and everything above 125 would be transformed to 255, or white. If you translate to grayscale as you generally will, you will get white and black.

Fig. Original frame

Fig. processed frame

All pre-processed images are given as input to the CNN classifier and each image is classified as per the training data. Classifier predicts the frame for the words in the trained data.

For training 1000 images are used for each class. The CNN is build using three 3D convolutional layer and one dense layer and it uses 10 epochs with 200 steps and 6500 validation steps at the time of training. The model is then stored in the ".h5" file. Then for the prediction classifier is created using the trained CNN model. Live processed frames are given to the classifier. If frame is a match and classified then it returns the word for that sign language action.

IV. NEURAL NETWORK EVALUATION

In this system, one of the well-known machine learning algorithms out there which is cast-off for image classification i.e. Convolutional Neural Network (or CNN). So basically, what is CNN – as we know it's a machine learning algorithm for machines to comprehend the features of image with a foresight and remember the features suggest whether the name of the new image fed to the classifier. We trained our model on words. After some initial testing, we found that using an initial base learning rate of few worked fairly well in fitting the training data - it provided a steady increase in accuracy and seemed to successfully converge. Once the improvements in the loss stagnated, we manually stopped the process and decreased the learning rate in order to try and increase our optimization of our loss function.

1. The model builds with training dataset up to 1/5 epoch and each epoch contains multiple keraslevel which calculate the loss and accuracy and iterate till model fits correctly.

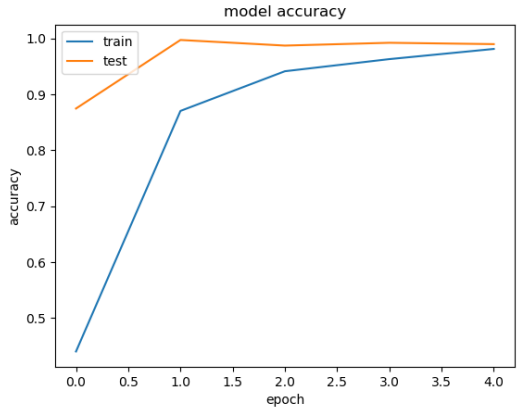


Fig. Model-1 Accuracy

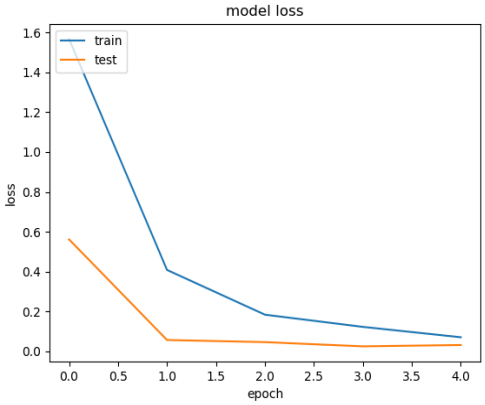


Fig. Model-1 Loss

2. The model builds with training dataset up to 1/7 epoch and each epoch contains multiple keras level and multiple steps each epoch which calculate the loss and accuracy and iterate till model fits correctly.

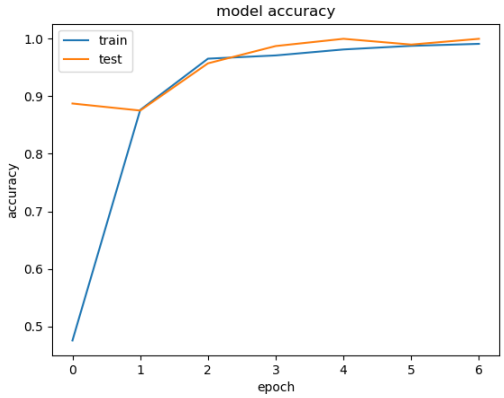


Fig. Model-2 Accuracy

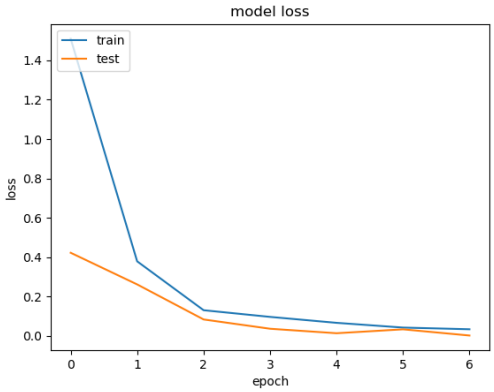


Fig. Model-2 Loss

3. The model builds with training dataset up to 1/10 epoch and each epoch contains multiple keras level and contains 500 step each epoch and 6500 validations which calculate the loss and accuracy and iterate till model fits correctly.

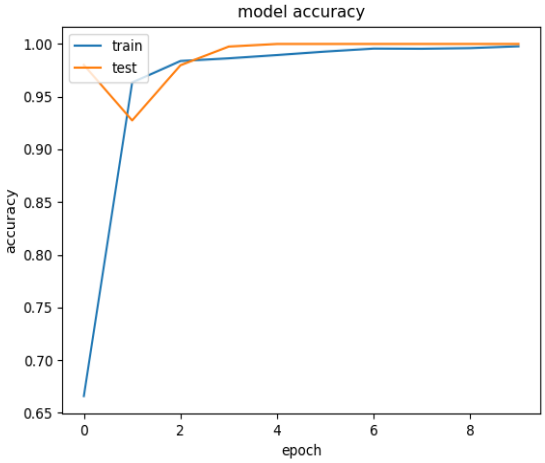


Fig. Model-3 Accuracy

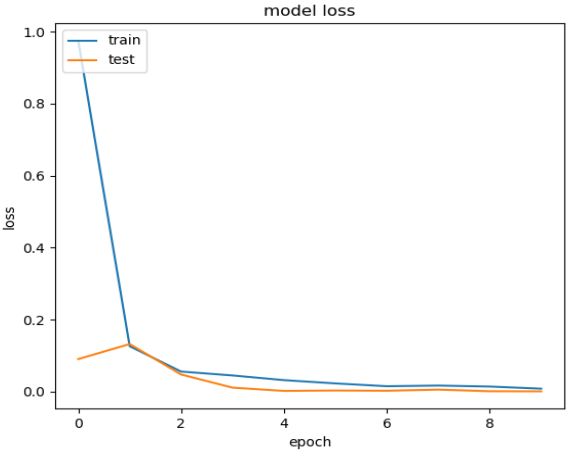


Fig. Model-3 Loss

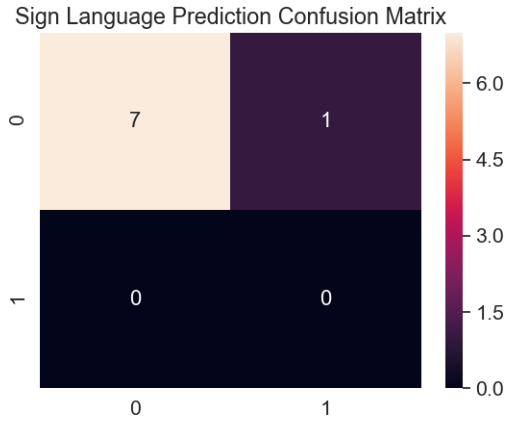
MODEL	ACCURACY	VALIDATION ACCURACY
1	0.8537	0.9316
2	0. 9246	0.9712
3	0.9529	0.9978

Loss and Accuracy:

losses were very noisy in the '1' and '2' models. Our space and time constraints initially required us to choose a less-than-optimal batch size value of 3, resulting in the noisy loss and accuracy we increased number of epochs and steps per epoch and validation steps to reduce our loss more smoothly and monotonically, in addition to more quickly converging on a validation accuracy.

Confusion Matrix:

The confusion matrices reveal that our accuracy suffers primarily due to the misclassification of words. Often the classifier gets confused between two or three similar words the confusion matrix for the word model reveals that with the exception of classifying "understand", it performed reasonably well.



V. EXPERIMENTAL RESULTS

The result of implemented proposed system are as follows:

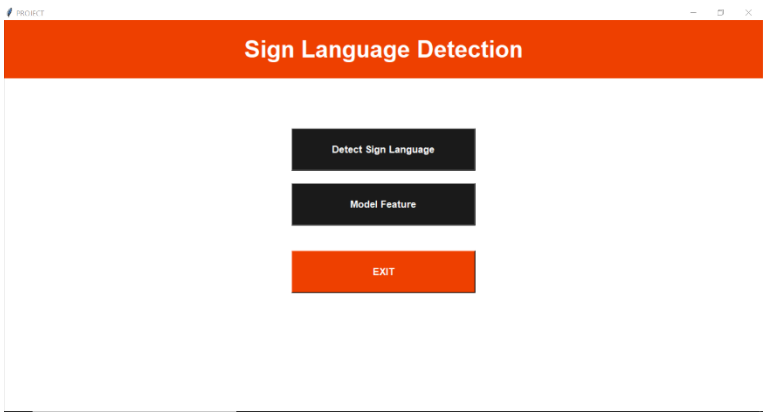


Fig. Main Page

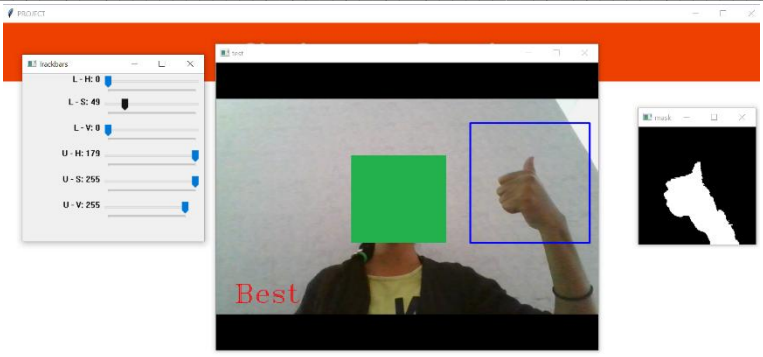


Fig. "Best" Sign Recognition

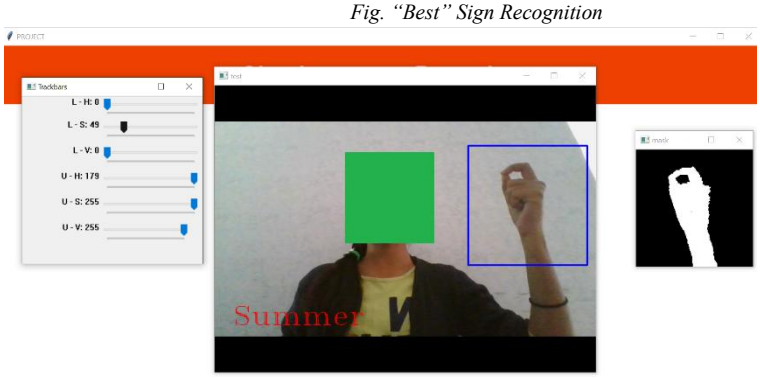


Fig. "Summer" Sign Recognition

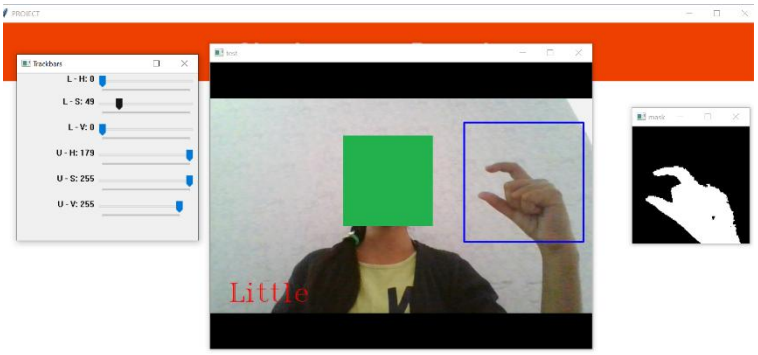


Fig. "Little" Sign Recognition

IV. CONCLUSION

The proposed system is developed for Indian English sign language recognition with Convolutional Neural Network using synthetic dataset. It detects the gesture and predict the textual output of the same by getting the input from the live video stream. The proposed system is able to determine the dynamic gestures as well as static gestures in real time environment without any need of specific colour gloves or specific colour background. The dataset used for this system consists of single handed actions having little movement. In future we can create the dataset consists of double handed actions as well as actions having more movement of hands. The efficiency achieved by the proposed CNN model is approximately 99% for the trained classes.

REFERENCES

[1] J. L. Raheja, A. Mishra, A. Chaudhary, "Indian sign language recognition using SVM", Pattern Recognition and Image Analysis, vol. 26, Issue. 2, pp. 434–441, June 2016.

[2] Y. I. Rokade, P. Jadav, "Indian sign language recognition system", International Journal of Engineering and Technology, vol. 9, pp. 189-196, July 2017.

[3] A. K. Saaho, "Indian sign language recognition using neural networks and kNN classifiers", Journal of Engineering and Applied Sciences, vol. 9, Issue. 8, pp. 1255-1259, August 2017.

[4] Z. Liang, S. Liao, B. Hu, "3D convolutional neural networks for dynamic sign language recognition", The Computer Journal, vol. 61, Issue. 11, pp. 1724-1736, November 2018.

[5] J. Singha, A. Roy, R. H. Laskar, "Dynamic hand gesture recognition using vision-based approach for human–computer interaction", Neural Computing and Applications, vol: 29, pp. 1129-1141, 2018.

[6] G. A. Rao, P. V. V. Kishore, "Selfie video based continuous Indian sign language recognition system", Ain Shams Engineering Journal, vol. 9, Issue. 4, pp. 1929–1939, 2018.

[7] K. Shenoy, T. Dastane, V. Rao, DevendraVyavaharkar, "Real-time Indian sign language(ISL) recognition", IEEE, ICCNT, October 2018 .

[8] X. Zhang, and X. Li, "Dynamic gesture recognition based on MEMP network", Future Internet, vol. 11, Issue. 4, 2019.

[9] M. R. Abid, E. M. Petriu, E. Amjadian, "Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar", IEEE transactions on instrumentations and measurements, vol. 64, Issue. 3, pp. 596-605, 2015.

[10] P. V. V Kishore, P. R. Kumar, E. K. Kumar and S. R. C. Kishore, "Video audio interface for recognizing gestures of indian sign language", International Journal of Image Processing (IJIP), vol. 5, Issue. 4, pp. 479-503, 2011.

[11] V. D. Edke, R. M. Kagalkar, "Review paper on video content analysis into text description", IJCA National Conference on Advances in Computing, Issue. 3, pp. 24-28, 2015.

[12] J. Farooq and M. B. Ali, "Real time hand gesture recognition for computer interaction", International Conference on Robotics and Emerging Allied Technologies in Engineering (ICREATE), pp. 22-24, April 2014.

[13] G. Plouffe and A. Cretu, "Static and dynamic hand gesture recognition system in depth data using dynamic time warping", IEEE Transactions on Instrumentation and Measurement, vol. 65, Issue. 2, pp. 305-316, February 2016.

[14] J. Ravikiran, K. Mahesh, S. Mahishi, Dheeraj R, S. Sudheender, N. V. Pujari, "Finger detection for sign recognition", Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. I, 2009.

[15] Z. A. Ansari, G. Harit, "Nearest neighbour classification of Indian sign language gestures using kinect camera", Indian Academy of Sciences, vol. 41, Issue. 2, pp. 161–182, February 2016.

[16] S. T. Hassan, J. A. Abolarinwa, C. O. Alenoghena, S. A. Bala, M. David, P. Enenche, "Intelligent sign language cognition using image processing techniques: A Case of Hausa Sign Language", ATBU Journal of Science, Technology and Education, vol. 6, Issue. 2, 2018.

[17] A. M. Jarman, S. Arshad, N. Alam, M. J. Islam, "An automated bengali sign language recognition system based on fingertip finder algorithm", International Journal of Electronics and Informatics, vol. 4, Issue. 1, July 2015.

[18] S. Konwar, S. Borah, T. Tuithung, "An american sign language detection system using HSV color model and edge detection", IEEE International Conference on Communication and Signal Processing, pp. 743-747, April 2014.